REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
Oct 94

3. REPORT TYPE AND DATES COVERED
Technical

4. TITLE AND SUBTITLE
Improvement of Inter-event Distance Tests of Randomness in Spatial Point Processes

5. FUNDING NUMBERS
DAAL03-92-G-0322

6. AUTHOR(S)
Linda Brant Collins, Anna Pluzhnikov, Michael L. Stein

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Chicago
Chicago, Illinois 60637

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 30167.7-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

A natural test of complete spatial randomness (CSR) for spatial point patterns based on the inter-event distance distribution has poor power against even obvious departures (Diggle 1983). We propose two modifications of the test statistic to improve its power. The first is a reweighting to put more emphasis on deviations from CSR at shorter distances. The second is to apply the theory of the projection of U-statistics to lower the variance of the test statistic under the null hypothesis. In examples where previous methods based on

CONT'D ON REVERSE SIDE

14. SUBJECT TERMS
listed on first page of report

15. NUMBER OF PAGES
18

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

inter-event distances failed to detect alternatives, we show that these modifications to the test statistic allow the detection of departures from CSR. Simulations of stationary and near stationary alternatives to CSR further illustrate the improvement in inter-event distance tests of CSR gained from weighting and projection. In addition, we propose two simple modifications to the graphical presentation of the empirical inter-event distance distribution to elucidate alternatives.

Improvement of Inter-event Distance Tests of Randomness in Spatial Point Processes *

by

Linda Brant Collins, Anna Pluzhnikov, and Michael L. Stein

TECHNICAL REPORT NO. 384

Accesion For	
NTIS CRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

Department of Statistics
The University of Chicago
5734 University Avenue
Chicago, Illinois 60637

October 21, 1994

* Computations and manuscript preparation were accomplished using computer facilities supported in part by the National Science Foundation grants DMS 89-05292, DMS 87-03942, and DMS 86-01732 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund. The first author was supported by Department of Education grant P200A10027. The second author was supported in part by Army Research Office grant DAAL03-92-G-0322. The third author was supported by National Science Foundation grant DMS 92-04504.

Improvement of Inter-event Distance Tests of Randomness in Spatial Point Processes

October 21, 1994

Key Words and phrases: edge effects, inter-event distance distribution, nearest-neighbor distribution, projection of U-statistics, tests of complete spatial randomness

AMS 1985 subject classifications: 60G55, 62M30

Abstract

A natural test of complete spatial randomness (CSR) for spatial point patterns based on the inter-event distance distribution has poor power against even obvious departures (Diggle 1983). We propose two modifications of the test statistic to improve its power. The first is a reweighting to put more emphasis on deviations from CSR at shorter distances. The second is to apply the theory of the projection of U-statistics to lower the variance of the test statistic under the null hypothesis. In examples where previous methods based on inter-event distances failed to detect alternatives, we show that these modifications to the test statistic allow the detection of departures from CSR. Simulations of stationary and near stationary alternatives to CSR further illustrate the improvement in inter-event distance tests of CSR gained from weighting and projection. In addition, we propose two simple modifications to the graphical presentation of the empirical inter-event distance distribution to elucidate alternatives.

1. Introduction

Analysis of a spatial point pattern often begins with a test of the hypothesis of complete spatial randomness (CSR). The null hypothesis of CSR asserts that conditional on the number of events in the spatial point pattern data set, the events are independent and uniformly distributed on the region of observation $A \subset \mathbb{R}^d$. We begin with a test of CSR since if CSR is not rejected, the process can be modeled as stationary Poisson and we proceed with estimation in this well understood setting. We can also view CSR as a central hypothesis separating "regular" (e.g. repulsion) processes on one side from "aggregated" (e.g. clustered) processes on the other (Diggle 1983). In this discussion, we propose improvements to the test of CSR based on the inter-event distance distribution $H(t)$ presented by Diggle (1983).

Under the null hypothesis of CSR, the inter-event distance distribution becomes

$$H_0(t) = \Pr\{|X_1 - X_2| \leq t\}$$

for X_1 and X_2 independently and uniformly distributed on the region A. Conditional on the number of events $N(A) = n$ in A where N is a spatial point process and $N(A) = N \cap A$, the empirical inter-event distance distribution function (EDF) is written

$$\hat{H}(t) = \frac{1}{n(n-1)} \sum_{i \neq j} I\{|x_i - x_j| \leq t\}$$

where x_i are the events in the observed spatial point pattern and $I\{\cdot\}$ is the indicator function. Under CSR, N is stationary Poisson and $\mathrm{E}[\hat{H}(t)] = H_0(t)$, so that deviations of $\hat{H}(t)$ from $H_0(t)$ can be used to test CSR. In this paper, we consider as examples data collected on square regions, for which Diggle (1983) gives an explicit expression for $H_0(t)$.

1.1. Improving Inter-event Distance EDF Statistics

Tests based on the empirical inter-event distance distribution have been criticized for inability to detect alternatives from CSR occurring at the small inter-event distances (Diggle 1983). In section 2, we describe a common inter-event distance EDF test and develop modifications in section 3 to improve the interpretability and power of the test. In particular, a reweighting of the statistic dampens the influence of large inter-event distances and results in the rejection of CSR in an example where obvious clustering at small distances is

undetected by the standard inter-event distance EDF test. Further, we apply the theory of the projection of U-statistics (Hájek 1968 and Serfling 1980) to $\hat{H}(t)$ to reduce its variance under CSR while only slightly changing the expected value of $\hat{H}(t)$ under stationary alternatives (Stein 1993, 1994). The projection provides a reduction from $O(n^{-1})$ to $O(n^{-2})$ in the asymptotic variance of the inter-event distance EDF statistic under the null (keeping t and the region of observation A fixed and letting $n \to \infty$). Section 3 provides evidence of improvement in examples where the significance level of the test decreases with these modifications to the statistic. Section 4 outlines the results of a simulation study comparing the power of the common test, its modifications, and a test based on the nearest-neighbor distribution.

1.2. Other Distance Method Tests of CSR

The current criticisms of tests based on the inter-event distance distribution encourage the use of alternative tests based on other point pattern descriptive distributions. For example, tests based on the nearest-neighbor distribution exhibit high power against point processes that depart from CSR at small inter-event distances (e.g. hard-core repulsion processes). The point-to-nearest-event distribution also leads to a test sensitive to departures from CSR occurring at the small inter-event distances. We prefer a distribution resulting in a test of CSR with power against a wider range of alternatives. The inter-event distance distribution and the closely related reduced second moment distribution hold information on pairwise interactions in the data throughout the range of inter-event distances. These two distributions have the potential to form the basis of tests with power against alternatives appearing at a wider range of inter-event distances.

Stein (1993) applies the projection theory of U-statistics to the empirical reduced second moment distribution to reduce the variance in the estimate under CSR. This should improve the power of reduced second moment EDF statistics for the testing problem. However, the edge-corrected estimates for the reduced second moment function are much more difficult to calculate than the inter-event distance EDF. In this paper, we apply the projection and reweighting to the empirical inter-event distance distribution. Implementation of these modifications is straightforward and the specific form for the projection of the inter-event distance is provided in section 2.3. The examples of section 3 and the simulations in section 4 compare the performance of the reweighted and/or projected statistics to the Diggle (1983) statistic and the analogous statistic for the nearest neighbor distribution. The increased clarity of presentation by modification of descriptive plots and increased power of tests based on projection and reweighting persuades us to reconsider the inter-event distribution as a useful tool in the detection of alternatives to CSR.

2. Inter-event Distance EDF Tests

2.1. Current Test Statistic

We begin by formally describing the Diggle (1983) approach to the test of CSR. A plot of $\hat{H}(t)$ versus $H_0(t)$ provides the basic data description tool. We would expect a linear plot for data consistent with CSR. For small values of t, large $\hat{H}(t)$ values compared to $H_0(t)$ indicate aggregation (e.g. clustering) and small $\hat{H}(t)$ values may indicate a process with regularity (e.g. repulsion). For large distances, large values of $\hat{H}(t)$ may indicate that events are sparse near the border of the region A while small values of $\hat{H}(t)$ can result from events dense near the border.

To compare $\hat{H}(t)$ to the variations we would expect under CSR, Diggle (1983) suggests performing s simulations of n points uniformly on A. A $100(1-\alpha)\%$ simulation envelope for $\hat{H}(t)$ is formed by calculating the EDF, $\hat{H}_i(t)$, for each simulation $i = 1, \ldots, s$ and recording the extreme values as described in Diggle (1983). The simulation envelope serves as the acceptance region for a test of CSR at any inter-event distance $t = t_0$.

Although the plot of $\hat{H}(t)$ versus $H_0(t)$ indicates both consistency of the data with CSR and the characterization of deviations from CSR, Diggle (1983) also considers a more formal test that avoids the subjectivity of fixing a test distance t_0 by averaging over t. Let

$$v_i = \int \left[\hat{H}_i(t) - H_0(t)\right]^2 dt,$$

for $i = 1, \ldots, s$, be the integrated squared difference between $\hat{H}_i(t)$ and $H_0(t)$. Let

$$v_0 = \int \left[\hat{H}(t) - H_0(t)\right]^2 dt$$

for the data. Rank the v_i for $i = 0, 1, \ldots, s$ in descending order. Under CSR, the $s+1$ possible rankings for v_0 are equally likely. If v_0 has rank k, this gives a test of CSR with attained significance $k/(s+1)$. This test is known to have low power and the examples in section 3 illustrate this problem. The lack of power in this test may result from large values of t contributing too much to the statistic v_0. That is, departures from CSR like regularity and aggregation show up at small distances, but for the remaining majority of t values, the process may closely conform to CSR.

In contrast, the nearest-neighbor EDF has higher power to detect departures from CSR occurring at small distances when using the integrated squared difference statistic, as occurs with the examples in section 3 below. Note, however, that the nearest-neighbor distribution increases quickly and is already near one at moderate t values. Hence, there is little difference between the nearest-neighbor distribution and its EDF at moderate and larger t values. The smaller distances contribute greatly to the integrated squared difference statistic. Rather than dismissing tests using the inter-event distance EDF because of low power against departures at small distances, we can consider alternative statistics that take advantage of the fact that the inter-event distance distribution holds information about the second-order properties of the process throughout the range of inter-event distances in the data.

2.2. Reweighted Test Statistic

Consider weighting the squared difference $\left[\hat{H}(t) - H_0(t)\right]^2$ at each value of t by a weight, $w(t)$, that gives more emphasis to the smaller t values. The test is carried out as before by ranking the $s+1$ values of

$$w_i = \int w(t) \left[\hat{H}_i(t) - H_0(t)\right]^2 dt,$$

for $i = 1, \ldots, s$, and

$$w_0 = \int w(t) \left[\hat{H}(t) - H_0(t)\right]^2 dt.$$

We will take $w(t) = H_0(t)^{-1}$ here. Note that it is certainly possible to consider alternate weighting functions and some are explored in the simulation study of repulsion processes in section 4.3.

2.3. Projection

The inter-event distance EDF is of the appropriate form to implement a variance reducing projection method introduced by Stein (1993). Stein (1993) suggested projection of the estimator of the reduced second moment measure, $K(t)$, for a spatial point process by examining a class of estimators of $K(t)$ possessing a certain unbiasedness constraint and then considering minimizing the mean squared error of the estimators under the null hypothesis that the process is Poisson. The method is general and can be readily applied to other estimators, such as the more easily calculated inter-event distance EDF.

First, note that conditional on $N(A) = n$ with N stationary Poisson, the standard, or unprojected estimator, $\hat{H}(t)$, is unbiased for $H_0(t)$. Considering t and the region A as fixed and letting $n \to \infty$ it can be

shown that

$$
\begin{aligned}
\operatorname{Var}\left[\hat{H}(t)\right] &= \frac{2}{n(n-1)}\operatorname{Var}\left[I\left\{|X_1 - X_2| \le t\right\}\right] \\
&\quad + \frac{4(n-2)}{n(n-1)}\operatorname{Cov}\left[I\left\{|X_1 - X_2| \le t\right\}, I\left\{|X_1 - X_3| \le t\right\}\right] \\
&= \frac{2}{n(n-1)}\operatorname{Var}\left[I\left\{|X_1 - X_2| \le t\right\}\right] + \frac{4(n-2)}{n(n-1)}\operatorname{Var}\left[\frac{\nu_d(b(X_1,t)\cap A)}{a}\right] \\
&= O(n^{-2}) + O(n^{-1})
\end{aligned}
\tag{2.3.1}
$$

where $b(x,t)$ is the ball of radius t centered at x and X_1, X_2, X_3 are independent and uniformly distributed on A. Here, $\nu_d(\cdot)$ indicates Lebesgue measure and $a = \nu_d(A)$. Figure 1 depicts the region $b(x,t)\cap A$. with A a square in $\mathbb{R}^2$.

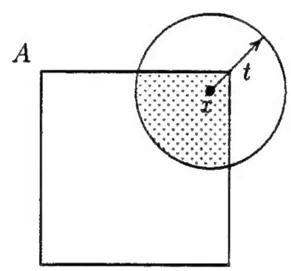


Figure 1. Example of $\nu_d(b(x,t)\cap A)$ (shaded region) for a square region $A \subset \mathbb{R}^d$ and ball of radius t centered at x.

If the region A has no edges (e.g. surface of a sphere or torus), then $\operatorname{Var}\left[a^{-1}\nu_d(b(X_1,t)\cap A)\right] = 0$ since $\nu_d(b(X_1,t)\cap A) = \nu_d(b(X_1,t)) = \nu_d(b(x,t))$ is constant. We gain pairwise independence for the random variables $I\{|X_i - X_j| \le t\}$ and $I\{|X_i - X_k| \le t\}$ on the sphere or torus so that the covariance term in (2.3.1) is zero (Silverman 1978). Thus, the asymptotic variance of the estimator becomes $O(n^{-2})$. Viewing $\hat{H}(t)$ as a U-statistic and projecting $\hat{H}(t)$ as in Serfling (1980, p. 188), results in a new estimator, $\hat{H}^P(t)$, that remains unbiased for $H_0(t)$ under CSR but has variance $O(n^{-2})$ regardless of any edge effects induced by the region A.

Applying the method outlined by Stein (1993), the projected inter-event distance EDF becomes

$$
\hat{H}^P(t) = \hat{H}(t) + 2H_0(t) - \frac{2}{an}\sum_{i=1}^{n}\nu_d\{b(x_i,t)\cap A\},
$$

with $\operatorname{Var}\{\hat{H}^P(t)\} = O(n^{-2})$ (Serfling 1980). Stein (1993) argues through asymptotics and simulation results that the bias introduced by projection is small for stationary N and the reduction in the mean squared

error for Poisson N is substantial for larger t. Here, the fact that the inter-event distance distribution holds information about pairwise interactions throughout the range of t values should improve the performance of projected inter-event distance EDF statistics over other tests that focus only on the small distances, such as those based on the nearest-neighbor EDF. In particular, since the nearest neighbor EDF is nearly constant for large values of t, any decrease in variability due to projection is likely to be small. Further, calculation of the projection of the nearest neighbor EDF will require a difficult integration to determine the conditional expectation of the nearest neighbor EDF given an event $x_i = x$, and this is not pursued here. Estimators of $K(t)$ can also form the basis of a test of CSR and be improved through projection as presented by Stein (1993, 1994), but these estimates involve considerably more calculation than the simpler inter-event distance EDF.

We recreate here the results Diggle (1983) obtained for three small examples by calculating $\hat{H}(t)$ and performing the Monte Carlo test as well as implementing the modifications of reweighting and projection described above. All calculations are carried out with $s = 99$ and 98% simulation envelopes.

3. Examples *3.1. Small Examples*

The data are: (1) the locations of 65 Japanese black pine saplings in a square of side 5.7 meters, (2) the locations of 62 redwood seedlings in a square of side 23 meters, and (3) the locations of 42 biological cell centers in a unit square. All data are as reported by Diggle (1983) from the references therein. For each example, a plot of the data (Figure 2(a)) normalized to the unit square is followed by a plot of $\hat{H}(t)$ versus $H_0(t)$ (Figure 2(b)). The plots in Figure 2(b) cover the entire range of inter-event distances in the unit square. Figure 2(c) provides the respective large scale plots focusing on the small distances t on a scale at which the alternatives of regularity and aggregation can be easily detected. In Figure 3(a), the rotated plots of $\hat{H}(t) - H_0(t)$ versus $H_0(t)$ clarify alternatives over the entire range of t even at the original small plot scale. In subsequent examples, we present only the rotated plots to evaluate results, although both modifications of the plot are useful in this situation where departures from CSR occur at the small distances. Figure 3(b) provides the plots of $\hat{H}^P(t) - H_0(t)$ versus $H_0(t)$ for the projected inter-event distance EDF. Since the simulations are performed under CSR, the resulting simulation envelopes for $\hat{H}^P(t)$ are tighter, as expected from the improvement in asymptotic variance under the null discussed in section 2.3.

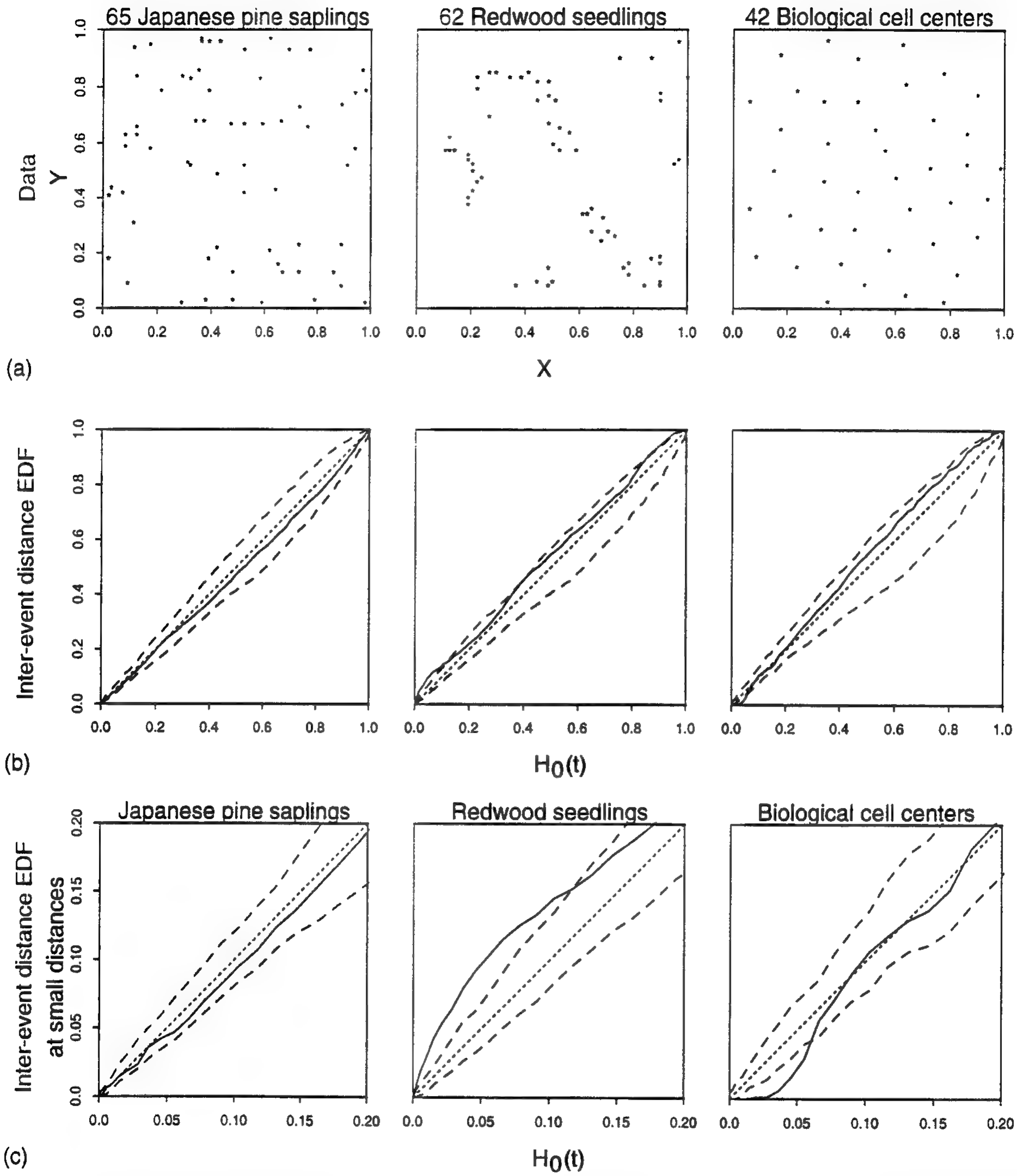


Figure 2. Data (a) and descriptive plots (b) and (c) (enlarged scale) of the inter-event distance EDF versus $H_0(t)$ (solid curve), 98% simulation envelopes from 99 simulations under CSR (dashed curves), and reference diagonal (dotted line) for three examples.

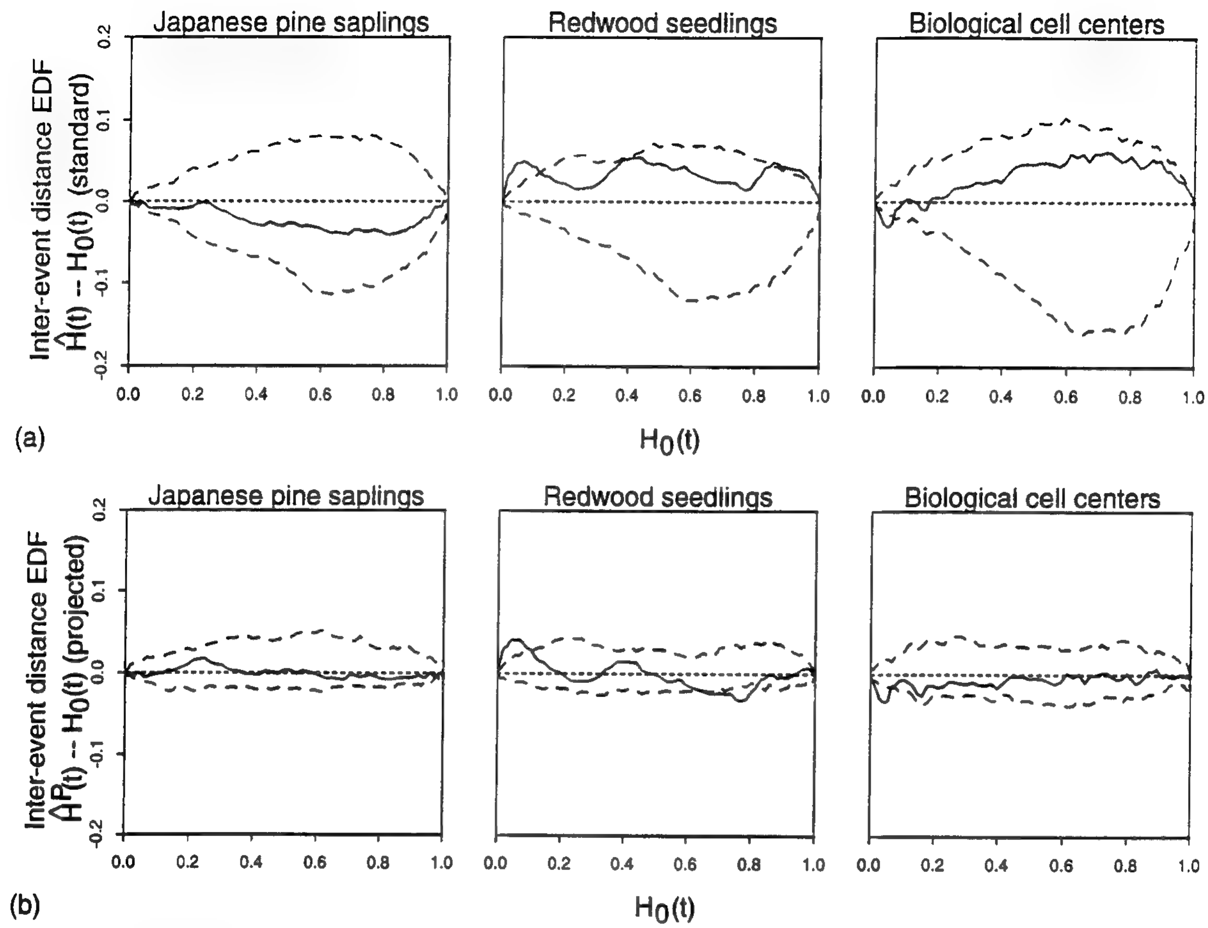


Figure 3. Descriptive plots of the difference between the inter-event distance EDF and $H_0(t)$ versus $H_0(t)$ (solid curve), 98% simulation envelopes from 99 simulations under CSR (dashed curves), and horizontal reference (dotted line) for three examples computed using (a) the standard (unprojected) inter-event distance EDF and (b) the projected EDF.

For the Japanese black pine saplings, we see that $\hat{H}(t)$ is clearly within the simulation envelope (see Figure 3). For the redwood seedlings, $\hat{H}(t)$ rises above the simulation envelope at small values of t indicating evidence of aggregation. This is expected since the data are clustered around locations of mature redwoods acting as parents in the seedling process. For the biological cell centers, $\hat{H}(t)$ falls below the simulation envelope at small distances, indicating regularity. Again, this is reasonable since the cell centers are restricted from coming in close proximity by the volumes of the cells around their centers. Note that $\hat{H}(t)$ lies within the simulation envelope for moderate and larger t-values for the biological cell centers. This implies that beyond the very small distances, the locations of the cell centers admit no particular structural

pattern. This is precisely the type of alternative that the integrated squared difference statistic has low power to detect.

Table 1 summarizes the results of the Monte Carlo tests of CSR using the integrated squared difference statistic for both the standard (unprojected) and projected EDF including the modification of reweighting the statistic.

Table 1: Comparison of attained significance levels for Monte Carlo tests of CSR based on 99 simulations using standard (unprojected) and projected inter-event distance EDF and nearest-neighbor EDF statistics for three examples.

	Statistic				
	v_0		w_0		z_0
Data Set Name	(standard)	(projected)	(standard, weighted)	(projected. weighted)	(nearest neighbor)
Japanese pine saplings	0.28	0.64	0.30	0.54	0.39
Redwood seedlings	0.12	0.04	0.02	0.01	0.01
Biological cell centers	0.26	0.23	0.16	0.01	0.01

Weighting helps detect the aggregation in the redwood seedlings with significance 0.02, but the regularity of the biological cell centers still escapes detection at the 5% error level with significance level 0.16. For both the redwood seedlings and biological cell centers, combining projecting and reweighting provides the clearest evidence against CSR. The attained significance levels for the analogous test of CSR based on an integrated squared difference statistic for the empirical nearest-neighbor distribution are provided. Here, we calculate the nearest neighbor EDF

$$\hat{G}(t) = \frac{1}{n}\sum_{i=1}^{n} I\{y_i \leq t\}$$

where y_i is the distance from event x_i to the nearest other event observed in A. The integrated squared difference statistic becomes

$$z_i = \int \left[\hat{G}_i(t) - G(t)\right]^2 dt,$$

for $i = 1, \ldots, s$ simulated data sets and

$$z_0 = \int \left[\hat{G}(t) - G(t)\right]^2 dt.$$

The nearest neighbor distribution, $G(t)$, depends on n and the region A, and is complicated due to edge effects. For this analysis, $G(t)$ is estimated under CSR by the average of $\hat{G}_j(t)$ for $j = 1, \ldots, 10{,}000$ simulations of CSR separate from the Monte Carlo test simulations. The nearest-neighbor EDF statistic clearly detects the departures of aggregation and repulsion from CSR at the small distances as expected.

3.2. Large Examples

After examining results for 11 additional examples treated by Diggle (1983), we come to the same conclusion for the test of CSR using any of the inter-event distance statistics discussed here (weighted, unweighted, unprojected, or projected) in all but two examples. Projection is necessary to reject CSR for the Lansing Woods red oak tree and white oak tree data from the collection of Lansing Woods data sets (Diggle 1983). Table 2 reports the attained significance levels for Monte Carlo tests of CSR based on 500 simulations using both standard (unprojected) and projected inter-event distance EDF statistics and the analogous integrated squared difference statistic for the nearest-neighbor EDF.

Table 2: Comparison of attained significance levels for Monte Carlo tests of CSR based on 500 simulations using standard (unprojected) and projected inter-event distance EDF and nearest-neighbor EDF statistics for two examples.

	Statistic				
	v_0		w_0		z_0
Lansing Woods Data Set Name	(standard)	(projected)	(standard, weighted)	(projected, weighted)	(nearest neighbor)
Red oak trees	0.363	0.002	0.228	0.002	0.002
White oak trees	0.214	0.012	0.176	0.002	0.032

The descriptive plots (Figure 4) show a departure from CSR at small distances for both the standard (unprojected) and projected EDFs, but CSR is rejected only with projection. The rotation of the descriptive plot becomes essential to its readability due to the tightened simulation envelope for the projected EDF. These examples where n is fairly large highlight the variance reduction realized through projection. In particular, Figure 4(a) demonstrates similarity in the standard (unprojected) and projected inter-event distance EDF statistic for the red oak data while the simulation envelope is much tighter with projection. The projected statistic clearly falls outside the simulation envelope even at the moderate distances. For the white oak data (Figure 4(b)), the projected EDF falls outside the simulation envelope at both small and large distances. We would expect the departures at large distances to contribute to a rejection of CSR. In fact, using both projection and reweighting results in rejection of CSR with attained significance of 0.002 (Table 2), while the analogous nearest-neighbor EDF test (focusing on small distances) has significance 0.032.

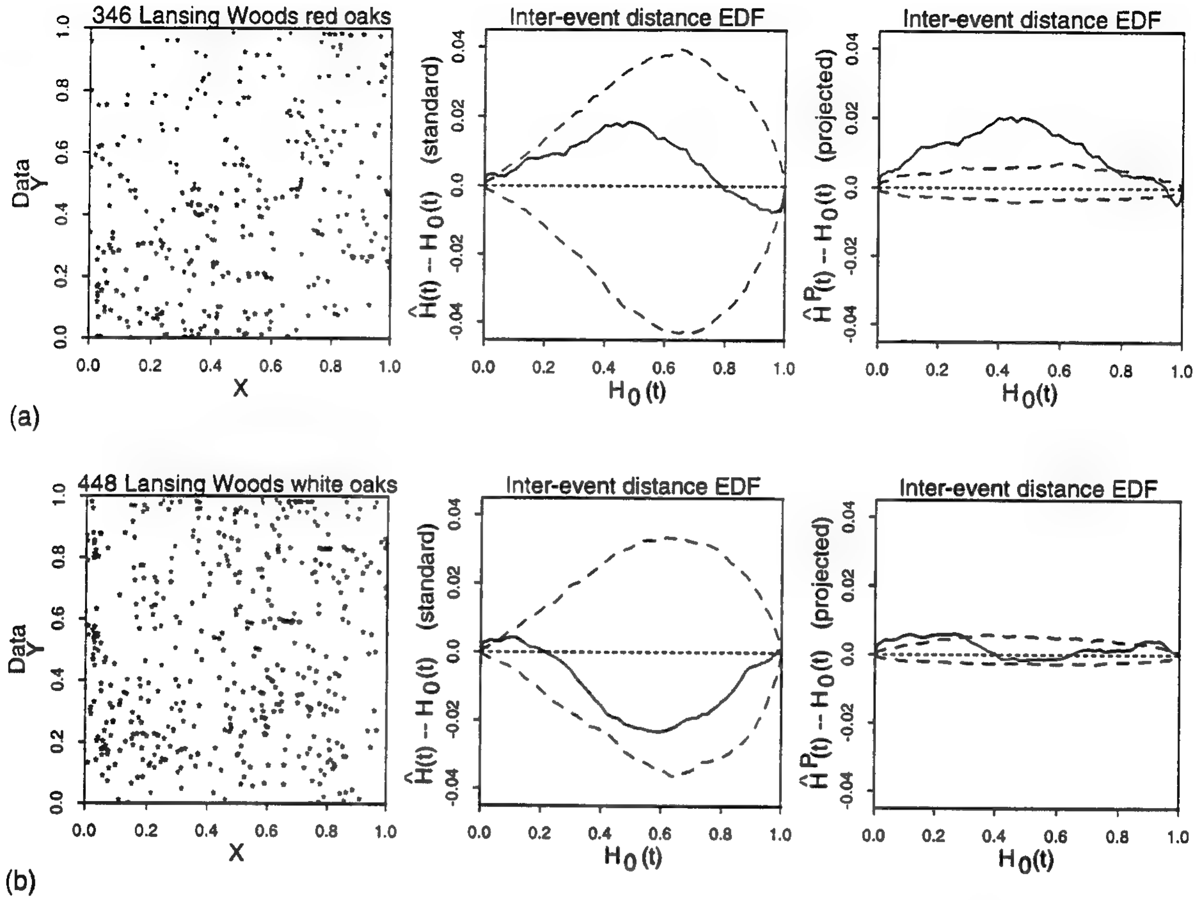


Figure 4. Data and descriptive plots of the difference between the inter-event distance EDF and $H_0(t)$ versus $H_0(t)$ for the standard (unprojected) and projected EDFs (solid curves), 98% simulation envelopes from 500 simulations under CSR (dashed curves), and horizontal reference (dotted line) for the Lansing Woods (a) red oak and (b) white oak data sets.

4. Simulation Study

As further evidence of the improvement of inter-event distance EDF tests through projection and reweighting of the integrated squared difference statistic, we present here a comparison of the performance of the various inter-event distance EDF tests and the nearest neighbor EDF test against several simulated departures from CSR. The simulated alternatives include a nonhomogeneous Poisson process with intensity that becomes "more stationary" with increasing n, the Neyman-Scott cluster process with varying cluster size and spread parameters, a "soft-core" repulsion process with varying repulsion distance distribution, and a stationary

process forced to have "too many" inter-event distances at a moderate distance. This last process demonstrates a departure from CSR only at moderate inter-event distances. For all simulated data sets with n events, the statistic from the data is compared to the values of the statistic for 1000 simulations of n events under CSR to determine significance at the 10% error level.

4.1. Nonhomogeneous Poisson Process

Conditional on the number of events n, consider the nonhomogeneous Poisson process with intensity $\lambda_n(x) = n\lambda + c\sqrt{n}\gamma(x)$ where $\gamma(x)$ is a fixed function of $x \in \mathbb{R}^d$, c a constant, and neither depending on n. The first term in this intensity function will dominate when n is large so that the process becomes "more stationary" as n increases. Figure 5 compares the proportion of rejections of CSR for the inter-event distance and nearest neighbor EDF tests at the 10% error level for 1000 simulated data sets with nonhomogeneous Poisson intensity $\lambda_n(x) = n + 20\sqrt{n}(x_1 + x_2)$ for $n = 30, 40, \ldots, 390, 400$ where x_1 and x_2 are the coordinates of x on the unit square $[0, 1]^2$.

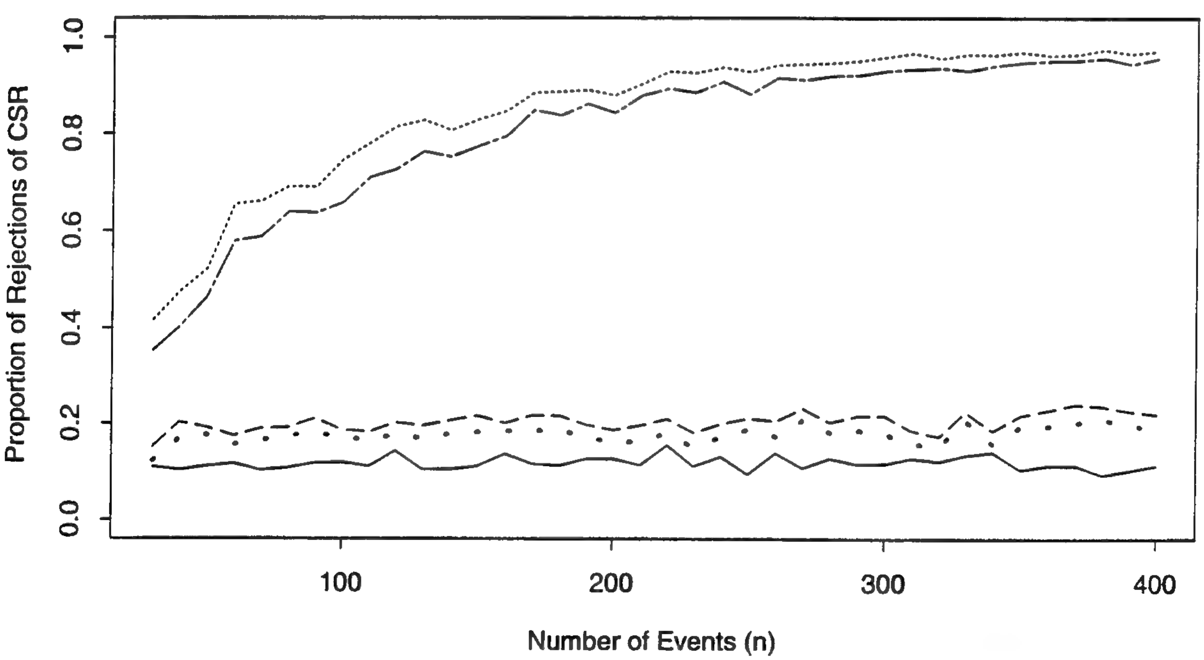
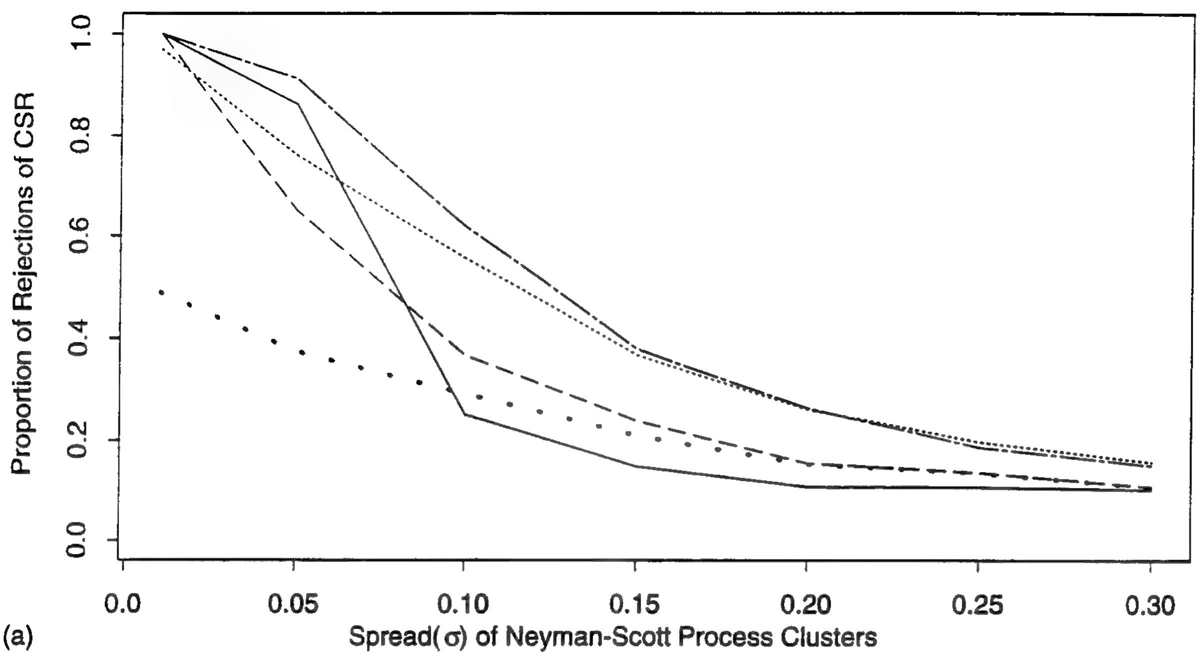


Figure 5. Proportion of rejections of CSR at the 10% error level for a nonhomogeneous Poisson alternative on the unit square using the integrated squared difference statistics based on the standard (unprojected) inter-event distance EDF (• • • • • •), the weighted statistic (— — — — —), the projected EDF (··············), the weighted statistic with projected EDF (—·—·—·—), and the nearest neighbor EDF (————————).

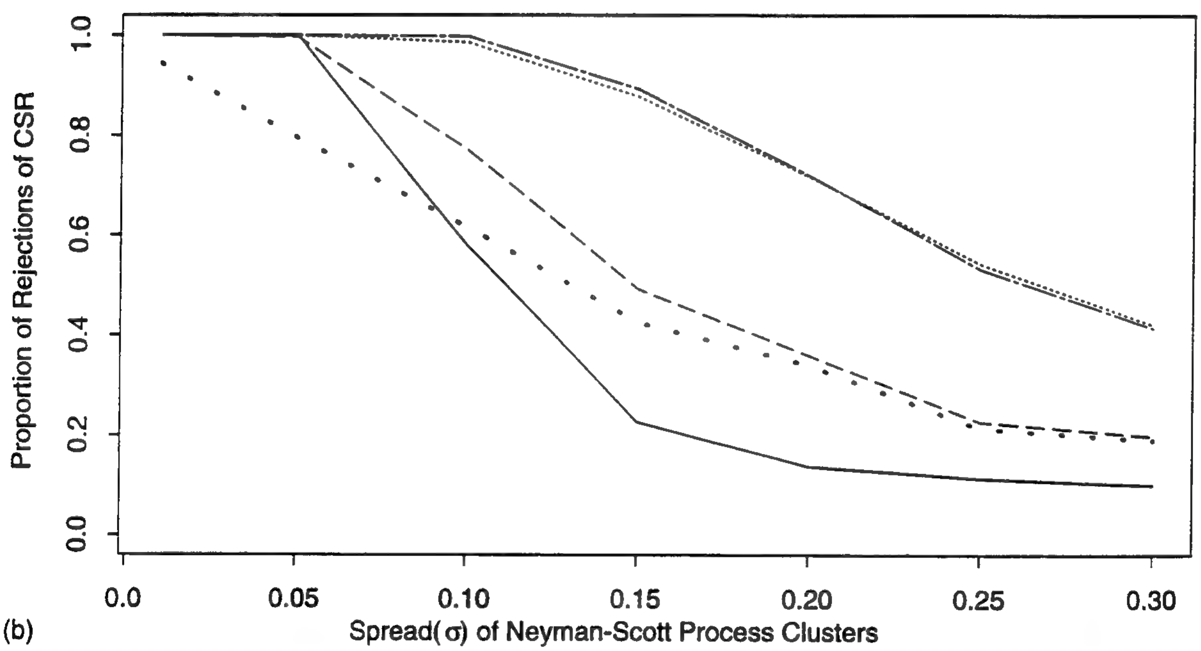
Projection clearly increases the power of the test against this alternative, especially for the more stationary realizations of this model when n is large.

4.2. Neyman-Scott Process

Consider the Poisson cluster process with parent events from a Poisson process with intensity ρ, Poisson(μ) number of offspring realized independently for each parent. and offspring independently and identically distributed about their parent according to a bivariate normal distribution with independent coordinates of mean zero and standard deviation σ. Note that this is equivalent to a Cox process (Diggle 1983, p. 58) and is also sometimes called a modified Thomas process (Stoyan, Kendall, and Mecke 1987, p. 144). This is a stationary process whose departure from CSR becomes more visible at small inter-event distances when the cluster size is large (μ large) or the spread of the cluster is small (σ small). We would expect a process with small cluster size and large spread to be difficult to discern from CSR using the nearest neighbor EDF test that concentrates on small distances. Figure 6 illustrates this behavior. This is a stationary alternative for which projection works well to detect the departure from CSR.



(a)



(b)

Figure 6. Proportion of rejections of CSR at the 10% error level for 1,000 realizations of a Neyman-Scott process on the unit square with parent intensity 20 and offspring intensity (a) 2 and (b) 10 using the integrated squared difference statistics based on the standard (unprojected) inter-event distance EDF (• • • • • •), the weighted statistic (— — — — —), the projected EDF (-------------), the weighted statistic with projected EDF (—·—·—), and the nearest neighbor EDF (————).

4.3. Soft-core Repulsion Process

For a repulsion process, we have simulated from the Matern-Bartlett Model as outlined in Cressie (1993). This model generalizes Matern's Model II to allow a distribution on the repulsion distance δ. We begin with a Poisson process of intensity ρ and then independently mark each event x with mark $m(x)$ from the uniform distribution on $[0, 1]$. The repulsion process results from a dependent thinning of the Poisson process driven by a positive nondecreasing and continuous function $f : [0,1] \rightarrow [0, \infty)$. An event x is deleted if there exists another event y such that both $m(y) < m(x)$ and $d(x, y) < f(m(y))$ hold. Figure 7 compares the power of the various EDF statistics against this alternative where the function f is taken to be

$$f(x) = \begin{cases} x, & \text{if } 0 \leq x \leq \delta \\ \delta, & \text{otherwise} \end{cases}$$

parameterized by δ.

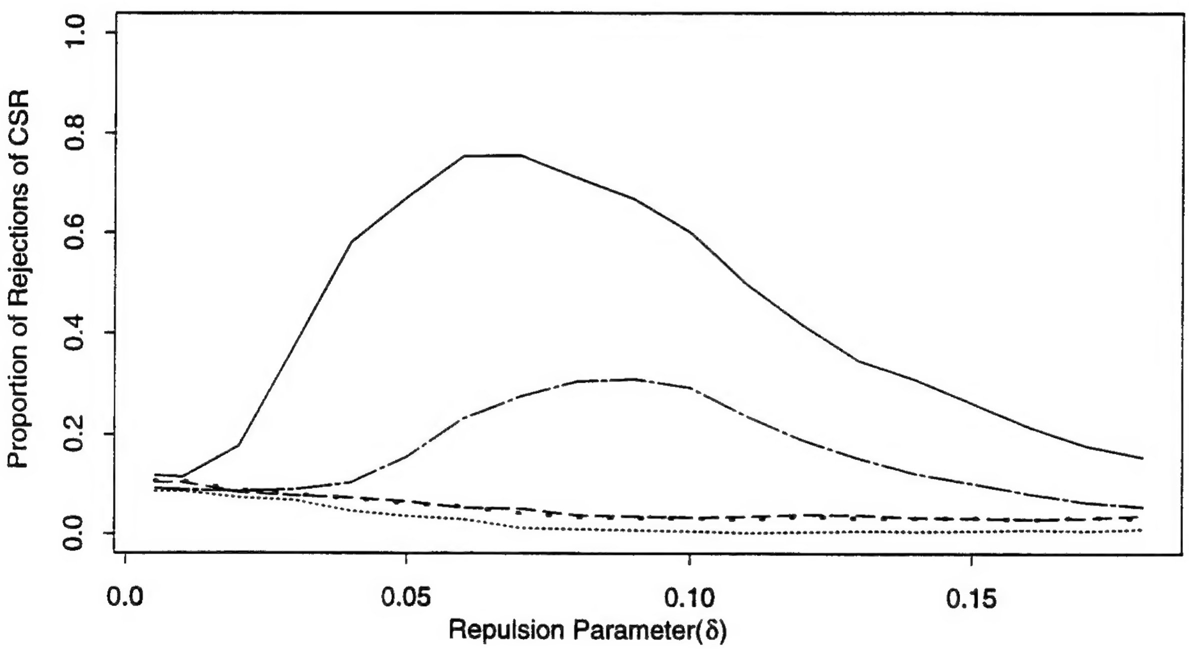


Figure 7. Proportion of rejections of CSR at the 10% error level for 1,000 realizations of a Matern-Bartlett Model repulsion process alternative using the integrated squared difference statistics based on the standard (unprojected) inter-event distance EDF (• • • • • •), the weighted statistic (— — — — —), the projected EDF (-------------), the weighted statistic with projected EDF (—·—·—), and the nearest neighbor EDF (————).

As the repulsion distance increases, the nearest neighbor gains power until the repulsion is so great that there is little data remaining in the thinned process. Neither weighting nor projection alone provide any

improvement over the standard inter-event distance EDF statistic. This is the behavior we also saw for the biological cell center example of a repulsion process in section 3. The nearest-neighbor distribution clearly has better power against this alternative. A stronger weighting of the small distances does somewhat improve the power of inter-event distance tests against this repulsion process. but the nearest-neighbor EDF still has the best power for this alternative.

Figure 8 compares the performance of several alternative weighting functions against the Matern-Bartlett repulsion process using the integrated squared difference statistic. The weight function, $w(t) = H_0(t)^{-3/2}$, gives more weight to the shorter distances than the suggested weight function, $w(t) = H_0(t)^{-1}$, and shows improved power against simulated repulsion processes. The stronger weighting of $w(t) = H_0(t)^{-2}$ improves power for departures from CSR at the very small distances (when δ is small), but quickly loses power as the repulsion distance increases.

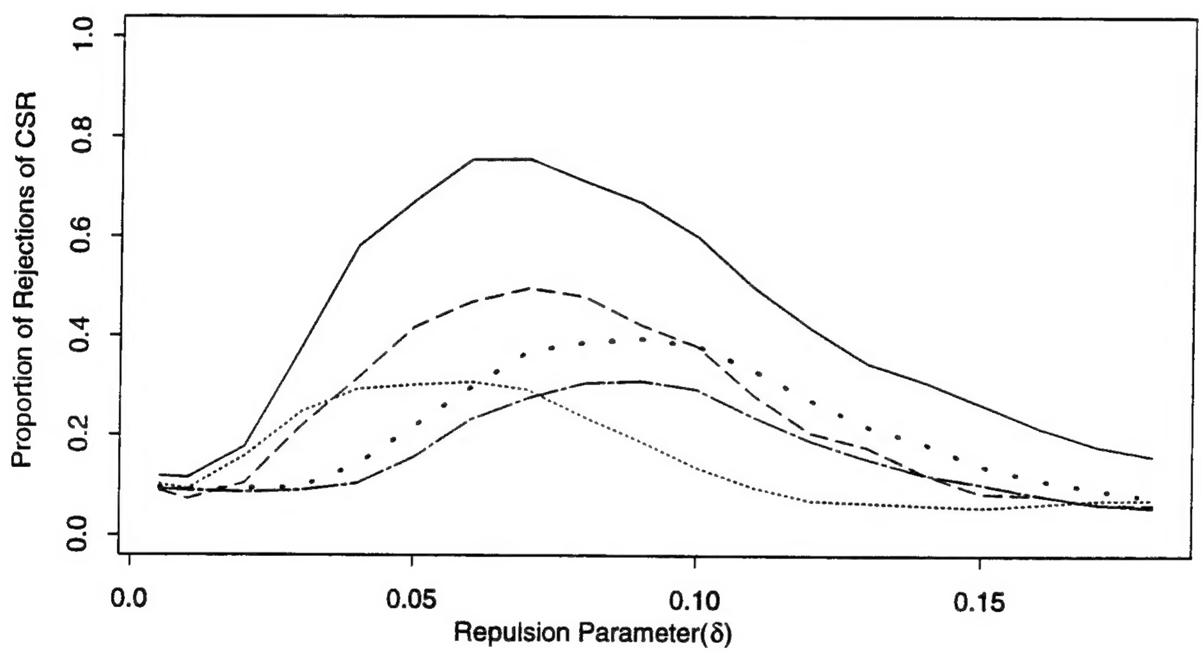


Figure 8. Proportion of rejections of CSR at the 10% error level for 1,000 realizations of a Matern-Bartlett Model repulsion process alternative using the integrated squared difference statistics based on the nearest neighbor EDF (————), the projected EDF with weighted statistic where $w(t) = H_0(t)^{-1}$ (—·—·—·), the projected EDF with weighted statistic where $w(t) = t^{-2}$ (· · · · · ·), the projected EDF with weighted statistic where $w(t) = H_0(t)^{-2}$ (··············), the projected EDF with weighted statistic where $w(t) = H_0(t)^{-3/2}$ (— — — — —).

4.4. *Stationary Process With Too Many Moderate Inter-event Distances*

Consider the following stationary process that departs from CSR by forcing too many inter-event distances at moderate distances. Start with a Poisson process of intensity γ. For each point in this process, place another point at a random distance r from the point in a random direction independent for each point. The random distance r is chosen from a distribution with expected value equal to a moderately large inter-event distance. When r is larger on average than the typical nearest neighbor distance expected, the nearest neighbor EDF test will have difficulty detecting any departure from CSR. This behavior is illustrated in Figure 9 where the random distance r is chosen from a N(0.1, 0.1/6) distribution. As n gets larger, the typical nearest neighbor distance expected under CSR gets smaller and the nearest neighbor EDF test quickly loses power against this alternative. Figure 9 compares the proportion of rejections of CSR for the various statistics for processes with intensity $\lambda = 2\gamma$ constructed by simulating a Poisson (γ) process on an extended region and adding one extra point at a random distance r for each point in the Poisson process. Points falling within the unit square are retained.

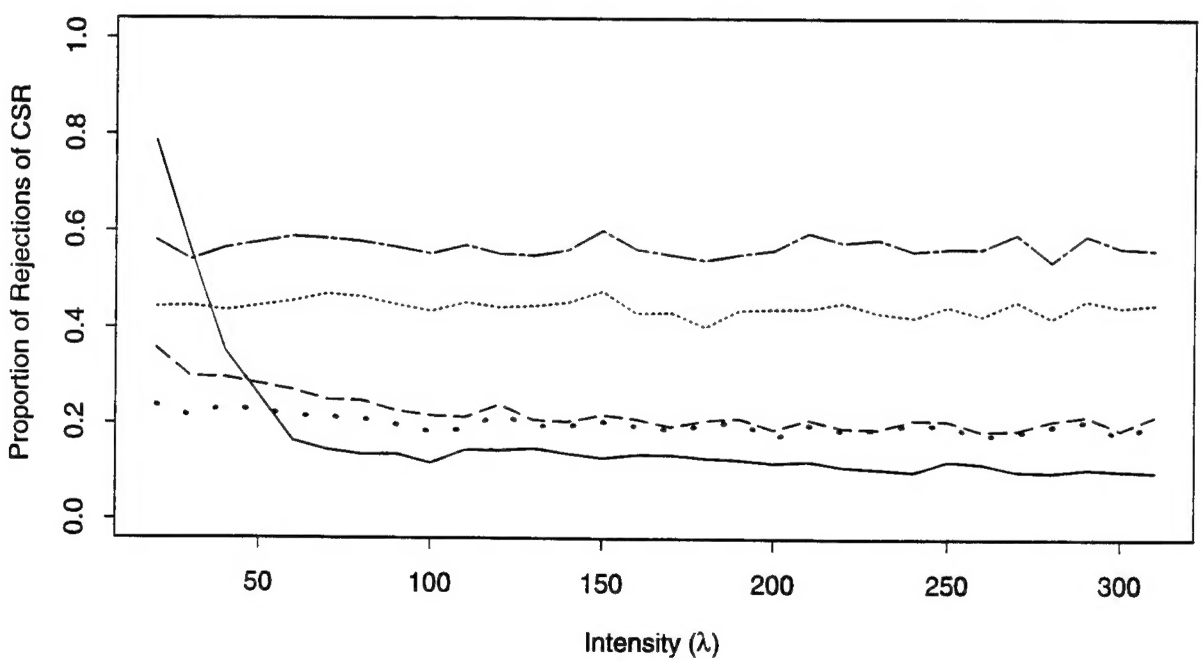


Figure 9. Proportion of rejections of CSR at the 10% error level for 1,000 realizations of a stationary alternative on the unit square forced to have too many inter-event distances at an average distance 0.1 using integrated squared difference statistics based on the standard (unprojected) inter-event distance EDF (• • • • • •), the weighted statistic (— — — — —), the projected EDF (·············), the weighted statistic with projected EDF (—·—·—·), and the nearest neighbor EDF (————————).

From Figure 9, we see that projection helps as expected, since projection reduces the mean squared error under CSR especially at the larger distances (Stein 1993). However, weighting alone does not help much since the departure from CSR occurs at moderate distances and weighting was introduced to decrease the influence of larger distances on the standard inter-event distance EDF statistic. This process illustrates a departure from CSR occurring only at moderate inter-event distances.

5. Conclusions

Inter-event distance EDF tests have been criticized for poor performance when departures from CSR occur at the small inter-event distances. The improvements to the EDF statistics from reweighting and projecting discussed in this paper argue that inter-event distance methods can have power against such alternatives. Further, the inter-event distance EDF, unlike the nearest-neighbor EDF, holds information on interactions in the data throughout the range of inter-event distances, giving inter-event distance methods the potential to detect a wider range of departures from CSR. Specifically, we can also hope to detect departures occurring only at moderate or large distances using inter-event distance methods.

References

Cressie, N. A. C. (1993). *Statistics for Spatial Data Patterns,* Revised Edition. New York: Wiley.

Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns.* London: Academic Press.

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternative. *Annals of Mathematical Statistics* **39**, 325-346.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability* **13**, 255-266.

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press, Cambridge.

Serfling, R. J. (1980). *Approximation Theorems in Mathematical Statistics.* Wiley, New York.

Silverman. B. W. (1978). Distances on circles, toruses and spheres. *Journal of Applied Probability* **15**, 136-143.

Stein, M. L. (1993). Asymptotically optimal estimation for the reduced second moment measure of point processes. *Biometrika* **80**, 443-449.

Stein, M. L. (1994). An approach to asymptotic inference for spatial point processes. In press, Statistica Sinica.

Stoyan, D.. Kendall, W. S., Mecke, J. (1987). *Stochastic Geometry and Its Applications.* Wiley, Chichester.